Serdica
Mathematical Journal
Сердика
Математическо списание

Serdica Math. J. **29** (2003), 11-32

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

A PVT-TYPE ALGORITHM FOR MINIMIZING A NONSMOOTH CONVEX FUNCTION

Li-Ping Pang, Zun-Quan Xia

Communicated by R. van Keer

Abstract. A general framework of the (parallel variable transformation) PVT-type algorithm, called the PVT-MYR algorithm, for minimizing a nonsmooth convex function is proposed, via the Moreau-Yosida regularization. As a particular scheme of this framework an ε-scheme is also presented. The global convergence of this algorithm is given under the assumptions of strong convexity of the objective function and an ε-descent condition determined by an ε-forced function. An appendix stating the proximal point algorithm is recalled in the last section.

1. Introduction.

A general framework of parallel computation for minimizing a nonlinear continuously differentiable function, called the parallel variable transformation (PVT) algorithm, was proposed by Fukushima [5], that is a synchro-paralleled structure. It is globally convergent at the linear rate under

2000 *Mathematics Subject Classification:* 90C25, 68W10, 49M37.

Key words: Parallel algorithm, synchronous parallel, convex minimization, Moreau-Yosida regularization, strong convexity, descent condition, forced function.

suitable conditions. The PVT algorithm can be regarded as an extension to the parallel variable distribution (PVD) algorithm, due to Ferris & Mangasarian [8] and developed by Solodov [17]. The PVT algorithm is also closely related to the parallel gradient distribution (PGD) algorithm due to Mangasarian [13]. In [19], Yamakawa and Fukushima studied performance of the PVT algorithm for unconstrained nonlinear optimization through numerical experiments. Also a number of other parallel algorithms were designed and developed for solving nonlinear optimization problems, see for instance, Han [9], Han and Lou [10], Bertsekas and Tsitsiklis [2], Liu and Tseng [12].

In this paper, a PVT-type algorithm, called the PVT-MYR algorithm, for minimizing a nonsmooth convex function, is proposed, which is constructed by converting an original objective function into a continuously differentiable function using the Moreau-Yosida regularization, due to Moreau [14] and Yosida [20].

The problem we are concerned with is of the form

$$\min_{x \in R^n} f(x), \tag{1.1}$$

where the objective function f defined on R^n is strongly convex, but not required to be smooth. A function f is said to be strongly convex if there exists a constant $c > 0$, called the modulus of strong convexity, such that

$$f(\alpha x + (1-\alpha)x') \le \alpha f(x) + (1-\alpha) f(x') - \frac{1}{2} c\alpha(1-\alpha)\|x - x'\|^2,$$

for all x, $x' \in R^n$ and $0 < \alpha < 1$, see [15].

Let F be the Moreau-Yosida regularization of f, $F : R^n \to R^1$, defined by

$$F(x) = \min_{z \in R^n} \left\{ f(z) + \frac{1}{2}\lambda^{-1}\|z - x\|^2 \right\}, \tag{1.2}$$

where λ is a positive parameter that will not be specified explicitly, following the way used in [16], and $\|\cdot\|$ denotes the Euclidean norm. It has been proved that F is finite convex, and the gradient $g = \nabla F$ is Lipschitzian. The unique minimizer $p(x)$ of (1.2) can be formulated in the form

$$p(x) = \arg\min_{z \in R^n} \left\{ f(z) + \frac{1}{2}\lambda^{-1}\|z - x\|^2 \right\}. \tag{1.3}$$

A point x is a solution of (1.1) iff it is a solution of the problem

$$\min_{x \in R^n} F(x), \tag{1.4}$$

see for instance, Hiriart-Urruty and Lemaréchal [11].

Some properties of the Moreau-Yosida regularization that will be used in this paper are listed below. Let f be convex, and ρ, c_1, c and c_2 be constants. The abbreviation 's. c.' stands for 'strongly convex', 's. m.' for 'strongly monotone'. The following properties can be referred to [11].

P1. F is Fréchet differentiable on R^n, g is Lipschitzian with constant λ^{-1}, and there exist $\rho > 0$ and $c_1 > 0$, such that $\|x - x^*\| \le \rho$ implies

$$F(x) - F(x^*) \le c_1 \|x - x^*\|^2, \tag{1.5}$$

where $x^* \in \operatorname{Arg\,min}_{x \in R^n} F(x)$.

P2. If f is s. c. with modulus c, then F is s. c. with modulus $c(c\lambda + 1)^{-1}$. If f is s. c., then there exist $\rho > 0$ and $c_2 > 0$ such that $\|x - x^*\| \le \rho$ implies that

$$\|g(x)\| \ge c_2 \|x - x^*\|. \tag{1.6}$$

P3. If f is s. c. with modulus c, then g is s. m. with modulus $c(c\lambda + 1)^{-1}$ on R^n, i. e.,

$$(g(x) - g(y))^T (x - y) \ge c(c\lambda + 1)^{-1} \|x - y\|^2, \tag{1.7}$$

for all x, $y \in R^n$.

Take an $\varepsilon > 0$. We can find an approximation, denoted by $p^a(x, \varepsilon) \in R^n$, to the unique minimizer $p(x)$ in (1.2) such that

$$\|p^a(x, \varepsilon) - p(x)\| \le \varepsilon \tag{1.8}$$

and

$$f(p^a(x, \varepsilon)) + \frac{1}{2}\lambda^{-1} \|p^a(x, \varepsilon) - x\|^2 \le F(x) + \varepsilon. \tag{1.9}$$

Let $F^a(x, \varepsilon)$, $g^a(x, \varepsilon)$ and $p^a(x, \varepsilon)$ be an $\varepsilon-$approximation to $F(x)$, to $g(x)$ and to $p(x)$, respectively, where the superscript, the little letter 'a', denotes the abbreviation for 'approximation', see for instance, Rauf and Fukushima [16], Fukushima [6], Correa and Lemaréchal [3] and Auslender [1], we define $F^a(x, \varepsilon)$ and $g^a(x, \varepsilon)$ to $F(x)$ and $g(x)$,

$$F^a(x, \varepsilon) = f(p^a(x, \varepsilon)) + \frac{1}{2}\lambda^{-1} \|p^a(x, \varepsilon) - x\|^2, \tag{1.10}$$

$$g^a(x,\varepsilon) = \lambda^{-1}(x - p^a(x,\varepsilon)). \tag{1.11}$$

It follows from (1.8)-(1.11) that $F^a(x,\varepsilon)$ and $g^a(x,\varepsilon)$ can be made arbitrarily close to the exact values of $F(x)$ and $g(x)$, respectively, in the process that ε tends to zero. This property given below can be found in Fukushima and Qi, [7].

P4. The following two inequalities are valid

$$F(x) \leq F^a(x,\varepsilon) \leq F(x) + \varepsilon, \tag{1.12}$$

$$\|g^a(x,\varepsilon) - g(x)\| \leq \sqrt{2\lambda^{-1}\varepsilon}. \tag{1.13}$$

It leads to the fact that $g^a(x,0) = g(x)$ and $F^a(x,0) = F(x)$.

The purpose of this paper is to present a PVT-type algorithm for solving (1.1), called the PVT-MYR algorithm presented in the next section, by combining the PVT algorithm and minimizing the Moreau-Yosida regularization F of f.

This paper is organized as follows. In Section 2, we present a PVT-type algorithm, the PVT-MYR algorithm, and establish its global convergence under some basic assumptions. In Section 3, we consider an ε-descent iteration condition for solving subproblems in the parallelization phase, and it is shown that the basic assumptions are satisfied under appropriate conditions on the transformations adopted. We present an ε-descent PVT-MYR algorithm and establish its global convergence. In Section 4, we establish a linear rate of convergence of the PVT-type (PVT-MYR) algorithm, presented in Section 2, under some additional assumptions. An appendix on a proximal point algorithm is given in the last section.

2. PVT-MYR algorithm.

We assume that the algorithm is implemented on p processors, where p is a positive integer. Each iteration of the algorithm consists of the parallelization phase and the synchronization phase. The former produces multiple candidate solutions for the next phase, using p processors, while the latter generates the next iterate point from the candidate solutions obtained in the parallelization phase.

For presenting the PVT-MYR algorithm some notations and assumptions are listed below.

BASIC NOTATIONS

p The number of parallel processors

m_l A positive integer such that $m_1 + m_2 + \cdots + m_p \geq n$

$E^{(k)}$ An $n \times (p+1)$ matrix whose columns consisting of $x^{(k)}$ and $A_l^{(k)} y_l^{(k)} + x^{(k)}$, $l = 1, \ldots, p$

$z^{(k)}$ $= (z_0^{(k)}, z_1^{(k)}, \ldots, z_p^{(k)})^T \in R^{p+1}$

$\varphi_l^{(k)}$ the auxiliary functions used in (2.2) and (2.4)

$\psi^{(k)}$ the auxiliary functions used in (2.3) and (2.5).

The following PVT steps are defined for solving the unconstrained smooth minimization problem

$$\text{(P)} \quad \min_{x \in R^n} f(x), \tag{2.1}$$

see [5].

PVT Algorithm: For unconstrained smooth minimization (P)

Step 0 `Initialization`

An initial point $x^{(0)} \in R^n$ is given and set $k = 0$.

Step 1 `Parallelization`

For each $l \in \{1, \ldots, p\}$, choose an $n \times m_l$ matrix $A_l^{(k)}$ and find an approximate solution $y_l^{(k)} \in R^{m_l}$ to the minimization problem

$$\min_{y_l \in R^{m_l}} \varphi_l^{(k)}(y_l) \equiv f(A_l^{(k)} y_l + x^{(k)}). \tag{2.2}$$

If $\nabla \varphi_l^{(k)}(0) = 0$, $l = 1, \cdots, p$, then stop. Otherwise, goto Step 2.

Step 2 `Synchronization`

Find an approximate solution $z^{(k)}$ to the minimization problem

$$\min_{z \in R^{p+1}} \psi^{(k)}(z) \equiv f(E^{(k)} z). \tag{2.3}$$

Set $x^{(k+1)} = E^{(k)} z^{(k)}$, $k = k + 1$. Loop at Step 1.

End of the PVT Algorithm

We now present a framework of the PVT-MYR algorithm for solving nonsmooth minimization problems.

PVT-MYR Algorithm: A general framework for nonsmooth minimization (P)

Step 0 `Initialization`

An initial point $x^{(0)} \in R^n$, constant $\varepsilon^* > 0$ and set $k = 0$.

Step 1 `Parallelization`

Step 1a `Initialization of the parallel step`

For each $l \in \{1, \ldots, p\}$, choose an $n \times m_l$ matrix $A_l^{(k)}$.

Step 1b `Compute the subproblem`

Find an approximate solution $y_l^{(k)} \in R^{m_l}$ to the minimization problem

$$\min_{y_l \in R^{m_l}} \varphi_l^{(k)}(y_l) \equiv F(A_l^{(k)} y_l + x^{(k)}), \tag{2.4}$$

where $F(A_l^{(k)} y_l + x^{(k)}) = \min_{z \in R^n} \{f(z) + 1/(2\lambda)\|z - A_l^{(k)} y_l - x^{(k)}\|^2\}$. If $\|\nabla \varphi_l^{(k)}(0)\| \leq \varepsilon^*$ for all $l \in \{1, \cdots, p\}$, then stop, otherwise, goto Step 2.

Step 2 `Synchronization`

Find an approximate solution $z^{(k)}$ to the minimization problem

$$\min_{z \in R^{p+1}} \psi^{(k)}(z) \equiv F(E^{(k)} z). \tag{2.5}$$

Set $x^{(k+1)} = E^{(k)} z^{(k)}$, $k = k + 1$ and loop at Step 1.

End of the PVT-MYR Algorithm

Remarks.

(i) Since the Moreau-Yosida regularization itself is defined through a minimization problem involving f, the exact calculation of the function F and its gradient g at point x is impossible in general. Therefore, in Step 1b we use approximation of these values instead of their exact values, such that $p^a(x, \varepsilon)$, $F^a(x, \varepsilon)$ and $g^a(x, \varepsilon)$ satisfy (1.8)-(1.11), respectively.

(ii) Note that approximate solutions to (2.4) computed in Step 1b are not required to be very accurate. In fact, for each l and k, we may only require that for some $y_l^{(k)}$ one has

$$\begin{aligned} \varphi_l^{(k)a}(y_l^{(k)}, \gamma\varepsilon^{(k)}) - \varphi_l^{(k)a}(0, \varepsilon^{(k)}) &= F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon^{(k)}) - F^a(x^{(k)}, \varepsilon^{(k)}) \\ &\leq -\eta \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 + \varepsilon^{(k)}, \end{aligned} \tag{2.6}$$

where $\eta > 0$, $\gamma \in (0, 1)$, $\varphi_l^{(k)a}(y_l^{(k)}, \gamma\varepsilon^{(k)}) = F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon^{(k)})$. The condition (2.6) is a key one for controlling the descent quantity of subproblem (2.4) in which the $\varepsilon^{(k)}-$slacked item in (2.6) is for implementation of this algorithm, PVT-MYR, and more details can be found in the next section.

(iii) The condition $A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)}) \le \varepsilon'$ is employed as the stopping criterion instead of the termination condition $\|\nabla \varphi_l^{(k)}(0)\| \le \varepsilon^*$. Then we have

$$\|A^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 = \sum_{l=1}^{p} \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 \le \varepsilon'^2 p.$$

By (A2) given below, see in (2.8), one has $\|g^a(x^{(k)}, \varepsilon^{(k)})\| \le \sqrt{p}\varepsilon'/\beta$. According to P4 and (A1) given below, we have $g(x^{(k)}) \le \sqrt{2\lambda^{-1}\gamma^k} + \sqrt{p}\varepsilon'/\beta$. If ε' and γ are small enough, then $\|g(x^{(k)})\|$ will be sufficiently small.

(iv) As for the synchronization phase, for each k, we may only require $x^{(k+1)}$ to satisfy

$$F^a(x^{(k+1)}, \varepsilon^{(k+1)}) \le \min_{1 \le l \le p} F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon^{(k)}). \tag{2.7}$$

In other words, $x^{(k+1)}$, $k = 1, \cdots,$ may be chosen as the ones that are determined by the $\varepsilon-$best candidates (in the sense of (2.7)), $A_{\bar{l}}^{(k)} y_{\bar{l}}^{(k)} + x^{(k)}$, $l = 1, 2, \ldots, p$.

Let $A^{(k)} = (A_1^{(k)}, \cdots, A_p^{(k)}) \in R^{n \times (m_1 + \cdots + m_p)}$. The following assumptions, marked by (A1)–(A4), and the definition, marked by (D4) are used for establishing the convergence of the PVT-MYR algorithm.

(A1) $\varepsilon^{(k)} \le \gamma\varepsilon^{(k-1)}$.

(A2) There exists a constant $\beta > 0$ independent of k such that

$$\|A^{(k)T} x\| \ge \beta\|x\|, \qquad \text{for all } x \in R^n. \tag{2.8}$$

(A3) There exists a constant $\delta_l > 0$ independent of k such that for all l, $\|A_l^{(k)}\| \le \delta_l$.

(A4) There exists a constant $\beta_l > 0$ independent of k such that $\|A_l^{(k)T} A_l^{(k)} y_l\| \ge \beta_l \|y_l\|$, for all $y_l \in R^{m_l}$.

(D1) We say that $\{d^{(k)} \mid k = 1, \ldots, \}$ satisfies a gradient relatedness condition if there exists a constant $\mu > 0$ such that the inequalities

$$\nabla F(x^{(k)})^T d^{(k)} \le -\mu \|\nabla F(x^{(k)})\| \cdot \|d^{(k)}\| < 0$$

are valid for all k.

Remarks. The following points should be mentioned

1. (A1) is made for implementation. (A1) $\Longrightarrow \sum_{k=1}^{\infty} \varepsilon^{(k)} < \infty$, when $\gamma \in (0,1)$.

2. The following two points can be referred to [5].

2a. (A2) $\Longleftrightarrow$ the sequence $\{A^{(k)}A^{(k)T}\}$ of $n \times n$ matrices is uniformly positive definite, i. e., there exists a constant $\beta' > 0$ independent of k such that

$$x^T A^{(k)} A^{(k)T} x \geq \beta' \|x\|^2, \qquad \text{for all} \quad x \in R^n.$$

2b. (A3) $\Longrightarrow \{A_l^{(k)}\}$ is uniformly bounded.

3. For implementation $\varepsilon^{(k)}$ is taken as the values for which the equality in (A1) is valid.

4. (A4) $\Longleftrightarrow$ the sequence $\{A_l^{(k)T} A_l^{(k)}\}$ of $m_l \times m_l$ matrices is uniformly positive definite.

5. The definition (D1) can be referred to [15].

Proposition 2.1 [10]. *If f is real-valued and convex over R^n, then x^* is the minimizer of $f(x)$ if and only if $g(x^*) = 0$ and $p(x^*) = x^*$.*

Lemma 2.1. *If $\lim_{k\to\infty} g^a(x^{(k)}, \varepsilon^{(k)}) = 0$, then $\lim_{k\to\infty} g(x^{(k)}) = 0$.*

Proof. By P4, we have

$$\|g^a(x^{(k)}, \varepsilon^{(k)}) - g(x^{(k)})\| \leq \sqrt{2\lambda^{-1}\varepsilon^{(k)}}. \tag{2.9}$$

According to (A1), it implies that $\lim_{k\to\infty} g(x^{(k)}) = 0$. $\square$

Lemma 2.2. *The following inequalities are valid*

$$\begin{aligned} F^a(x^{(k+1)}, \varepsilon^{(k+1)}) - F^a(x^{(k)}, \varepsilon^{(k)}) - \varepsilon^{(k+1)} &\leq F(x^{(k+1)}) - F(x^{(k)}) \\ &\leq F^a(x^{(k+1)}, \varepsilon^{(k+1)}) - F^a(x^{(k)}, \varepsilon^{(k)}) + \varepsilon^{(k)}. \end{aligned}$$

Proof. By P4, we obtain

$$F(x^{(k+1)}) \leq F^a(x^{(k+1)}, \varepsilon^{(k+1)}) \leq F(x^{(k+1)}) + \varepsilon^{(k+1)},$$

$$F(x^{(k)}) \leq F^a(x^{(k)}, \varepsilon^{(k)}) \leq F(x^{(k)}) + \varepsilon^{(k)}.$$

Combining the two inequalities given above, we have

$$F^a(x^{(k+1)},\varepsilon^{(k+1)}) - F^a(x^{(k)},\varepsilon^{(k)}) - \varepsilon^{(k+1)} \le F(x^{(k+1)}) - F(x^{(k)})$$
$$\le F^a(x^{(k+1)},\varepsilon^{(k+1)}) - F^a(x^{(k)},\varepsilon^{(k)}) + \varepsilon^{(k)}. \qquad \square$$

The following theorem is one of the main results on the convergence analysis of the PVT-MYR algorithm.

Theorem 2.1. *If the following conditions are satisfied*

a. *The objective function f is strongly convex;*

b. *At each iteration of the PVT-MYR algorithm,* (2.6) *and* (2.7) *are satisfied in Step* 1b *and Step* 2*, respectively;*

c. (A1) *and* (A2) *are satisfied,*

then any cluster of iterate points (estimates) generated by the PVT-MYR algorithm is the minimal solution of (P).

Proof. By (2.6) and (2.7), one has

$$F^a(x^{(k+1)},\varepsilon^{(k+1)}) - F^a(x^{(k)},\varepsilon^{(k)}) \le -\eta \|A_l^{(k)T} g^a(x^{(k)},\varepsilon^{(k)})\|^2 + \varepsilon^{(k)}.$$

Then, one has from Lemma 2.2 that

$$\begin{aligned} F(x^{(k+1)}) - F(x^{(k)}) &\le F^a(x^{(k+1)},\varepsilon^{(k+1)}) - F^a(x^{(k)},\varepsilon^{(k)}) + \varepsilon^{(k)} \\ &\le -\eta \|A_l^{(k)T} g^a(x^{(k)},\varepsilon^{(k)})\|^2 + 2\varepsilon^{(k)} \\ &\le -\eta \|A_l^{(k)T} g^a(x^{(k)},\varepsilon^{(k)})\|^2 + 2\gamma^k \varepsilon^{(0)} \end{aligned} \tag{2.10}$$

for $\gamma \in (0,1)$. For proceeding by contradiction, suppose that $\liminf_{k\to\infty} \|A_l^{(k)T} g^a(x^{(k)},\varepsilon^{(k)})\| = \sigma > 0$, for some l. Then there exists an infinite index set K, such that for $k \in K$ one has

$$\|A_l^{(k)T} g^a(x^{(k)},\varepsilon^{(k)})\| > \frac{1}{2}\sigma.$$

Thus, for $\gamma \in (0,1)$ small enough, one has

$$F(x^{(k)}) - F(x^{(k+1)}) > 0.$$

This implies that $\{F(x^{(k)})\}$ is decreasing. Since f is strongly convex, it follows by P2 that F is strongly convex and bounded below. This leads to

$$\lim_{k\to\infty} F(x^{(k)}) = F^*, \qquad k \in K,$$

where F^* is some real number. Adding (2.10) with respect to k, one has

$$\begin{aligned} F(x^{(0)}) - F^* &\geq \eta \sum_{k \in K} \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 - 2\varepsilon^{(0)} \sum_{k \in K} \gamma^{(k)} \\ &\geq \frac{1}{2} \eta \sum_{k \in K} \sigma - 2\varepsilon^{(0)} \gamma (1 - \gamma)^{-1}. \end{aligned}$$

It leads to

$$\sum_{k \in K} \sigma \leq 2\eta^{-1}[F(x^{(0)}) - F^* + 2\varepsilon^{(0)} \gamma (1 - \gamma)^{-1}].$$

This is impossible since K is infinite, and $\sigma > 0$ and F^* are finite. Therefore, we obtain

$$\liminf_{k \to \infty} \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\| = 0, \quad \text{for all } l. \tag{2.11}$$

On the other hand, noticing that $A^{(k)} = (A_1^{(k)}, \ldots, A_p^{(k)})$, we obtain from (A2) that

$$\begin{aligned} \sum_{l=1}^{p} \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 &= \|A^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 \\ &\geq \beta \|g^a(x^{(k)}, \varepsilon^{(k)})\|^2. \end{aligned}$$

This implies that

$$\lim_{\substack{k \to \infty \\ k \in K}} g^a(x^{(k)}, \varepsilon^{(k)}) = 0, \tag{2.12}$$

according to (2.11). It follows from Lemma 2.1, and (2.12) that

$$\lim_{\substack{k \to \infty \\ k \in K}} g(x^{(k)}) = 0,$$

which implies that every cluster of $\{x^{(k)}\}$ is the unique solution of (P). $\square$

3. An ε-descent direction for solving subproblems. It can be seen from the last section that it is not necessary to provide accurate solutions of (2.4) at each iteration when the PVT-MYR algorithm is performed, more specifically, at iteration k, it is sufficient to find a $y_l^{(k)}$ such that (2.6) is satisfied when we minimize each lth auxiliary function $\varphi_l^{(k)}$ with respect to y_l, starting with the origin, $y_l = 0$. Suppose that $y_l^{(k)}$ is determined by

$$y_l^{(k)} = \alpha_l^{(k)} d_l^{(k)}. \tag{3.1}$$

A direction $d_l^{(k)}$ is said to be ε-descent if the following gradient-relatedness condition of the direction $A_l^{(k)} d_l^{(k)}$ in the sense of Ortega and Rheinboldt [15] is satisfied

$$[A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})]^T d_l^{(k)} \leq -\mu_0 \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|d_l^{(k)}\| < 0, \tag{3.2}$$

with choices of $\mu_0 > 0$ and $\alpha_l^{(k)} > 0$ obeying the Armijo rule

$$\begin{aligned} \varphi_l^{(k)a}(y_l^{(k)}, \gamma\varepsilon^{(k)}) - \varphi_l^{(k)a}(0, \varepsilon^{(k)}) &= F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon_k) - F^a(x^{(k)}, \varepsilon^{(k)}) \\ &\leq \mu_1 \alpha_l^{(k)} [A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})]^T d_l^{(k)} + \varepsilon^{(k)}, \end{aligned} \tag{3.3}$$

in which parameters are determined or defined by the following conditions
$\mu_1 \in (0,1)$,
$\alpha_l^{(k)} = \beta_l^{m_k}$,
m_k is the smallest nonnegative integer number satisfying (3.3),
$\varepsilon^{(k)} \leq \beta^{(k)} \|y_l^{(k)}\|^2, \beta^{(k)} \to 0, k \to \infty$.

Remarks.

i. The difference between the statement here and the one by Ortega & Rheinboldt (1970), [15], consists in that condition (3.2) is replaced by

$$[A_l^{(k)T} g(x^{(k)})]^T d_l^{(k)} \leq -\mu_0 \|A_l^{(k)T} g(x^{(k)})\| \cdot \|d_l^{(k)}\| < 0,$$

and condition (3.3) is replaced by

$$F(A_l^{(k)} y_l^{(k)} + x^{(k)}) - F(x^{(k)}) \leq \mu_1 \alpha_l^{(k)} [A_l^{(k)T} g(x^{(k)})]^T d_l^{(k)},$$

see [15].

ii. The procedure of Armijo type described in (3.3) has been used in [7], [4] and [18].

Lemma 3.1 [16]. *Let $\varepsilon_x, \varepsilon_y > 0$ be arbitrary, and $\varepsilon = \max(\varepsilon_x, \varepsilon_y)$. If f is strongly convex with modulus c, then the following inequalities hold for all $x, y \in R^n$*

$$\langle g^a(x, \varepsilon_x) - g^a(y, \varepsilon_y), x - y\rangle \geq c/(c\lambda + 1)\|x - y\|^2 - \sqrt{8\varepsilon/\lambda}\|x - y\|. \tag{3.4}$$

Lemma 3.2 [16]. *If* (3.4) *is satisfied, then there exist positive constants m and M, and a positive integer k_0 such that*

$$\begin{aligned} \langle y_k, s_k\rangle/\|s_k\|^2 &\geq m, \\ \|y_k\|^2/\langle y_k, s_k\rangle &\leq M, \end{aligned} \tag{3.5}$$

for all $k \geq k_0$.

Lemma 3.3 [16]. *If* (3.5) *is satisfied, then there exist positive constants $\widetilde{\beta}$ and $\overline{\beta}$ such that the inequalities*

$$d^{(k)T}B^{(k)}d^{(k)} \geq \widetilde{\beta}\|B^{(k)}d^{(k)}\| \cdot \|d^{(k)}\|,$$

$$\|B^{(k)}d^{(k)}\| \leq \overline{\beta}\|d^{(k)}\|$$

are satisfied for infinitely many k, where $B^{(k)}$ is updated by the BFGS formula

$$B^{(k+1)} = B^{(k)} - \frac{B^{(k)}s^{(k)}s^{(k)^T}B^{(k)}}{s^{(k)T}B^{(k)}s^{(k)}} + \frac{y^{(k)}y^{(k)T}}{s^{(k)T}y^{(k)}},$$

where $s^{(k)} = x^{(k+1)} - x^{(k)}$, $y^{(k)} = g^a(x^{(k+1)}, \varepsilon^{(k+1)}) - g^a(x^{(k)}, \varepsilon^{(k)})$ and $0 < \varepsilon^{(k+1)} < \varepsilon^{(k)}$.

Lemma 3.4. *If $d^{(k)}$ is computed by $d^{(k)} = -B^{(k)-1}g^a(x^{(k)}, \varepsilon^{(k)})$, then $d^{(k)}$, $k \in K$, satisfy a gradient-relatedness condition, i.e.,*

$$g^a(x^{(k)}, \varepsilon^{(k)})^T d^{(k)} \leq -\widetilde{\beta}\|g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|d^{(k)}\| < 0, \tag{3.6}$$

where K is an infinite set.

Proof. By calculating, we have

$$\begin{aligned} g^a(x^{(k)}, \varepsilon^{(k)})^T d^{(k)} &= -g^a(x^{(k)}, \varepsilon^{(k)})^T B^{(k)-1} g^a(x^{(k)}, \varepsilon^{(k)}) \\ &= -[B^{(k)-1}g^a(x^{(k)}, \varepsilon^{(k)})]^T B^{(k)} [B^{(k)-1}g^a(x^{(k)}, \varepsilon^{(k)})] \\ &= -d^{(k)T}B^{(k)}d^{(k)}. \end{aligned} \tag{3.7}$$

It follows from Lemma 3.3 that

$$d^{(k)T}B^{(k)}d^{(k)} \geq \widetilde{\beta}\|B^{(k)}d^{(k)}\| \cdot \|d^{(k)}\| = \widetilde{\beta}\|g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|d^{(k)}\|. \tag{3.8}$$

Combining (3.7) and (3.8), we obtain

$$g^a(x^{(k)}, \varepsilon^{(k)})^T d^{(k)} \leq -\widetilde{\beta}\|g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|d^{(k)}\|. \qquad \square$$

Lemma 3.5. *If a direction* $d_l^{(k)}$ *satisfies the equation*

$$A_l^{(k)} d_l^{(k)} = d^{(k)},$$

and (A3) *and* (A4) *are satisfied, then the inequality given by* (3.2) *is valid.*

Proof. From Lemma 3.4, we have

$$\begin{aligned} & (A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)}))^T d_l^{(k)} \\ \leq \; & -\widetilde{\beta}\|g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|A_l^{(k)} d_l^{(k)}\| \\ \leq \; & -\widetilde{\beta}/\delta_l^2 \|A_l^{(k)}\| \cdot \|g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|A_l^{(k)}\| \cdot \|A_l^{(k)} d_l^{(k)}\| && \text{from (A3)} \\ \leq \; & -\widetilde{\beta}/\delta_l^2 \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|A_l^{(k)T} A_l^{(k)} d_l^{(k)}\| \\ \leq \; & -\widetilde{\beta}\beta_l/\delta_l^2 \|A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\| \cdot \|d_l^{(k)}\|, && \text{from (A4)} \end{aligned} \tag{3.9}$$

where the third inequality comes from Cauchy-Schwartz inequality. Setting

$$\mu_0 = \min_{1 \leq l \leq p} \{\widetilde{\beta}\beta_l/\delta_l^2\},$$

then the inequality (3.2) is valid. $\square$

The lemma given above shows that there exists a direction satisfying (3.2). For convenience, the following notations are given

$$\begin{aligned} \widetilde{\mu} \; &:= 1 - \mu_1 \\ \nabla\varphi_l^{(k)} \; &:= A_l^{(k)T} g(x^{(k)}) \\ \nabla^a\varphi_l^{(k)} \; &:= A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)}). \end{aligned}$$

Lemma 3.6. *Suppose* $y_l^{(k)}$ *is determined by* (3.1) *with* $d_l^{(k)}$ *and* $\alpha_l^{(k)}$ *satisfying* (3.2) *and* (3.3). *If* (A3) *is satisfied, then for each* l *one has that* (2.6) *holds.*

P r o o f. From (3.3), one has that if $\widetilde{\alpha}_l^{(k)} = 2\alpha_l^{(k)}$, then the line search must be failing, i. e.,

$$\begin{aligned}\varphi_l^{(k)a}(\widetilde{y}_l^{(k)}, \gamma\varepsilon^{(k)}) - \varphi_l^{(k)a}(0, \varepsilon^{(k)}) &= F^a(A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}, \gamma\varepsilon_k) - F^a(x^{(k)}, \varepsilon^{(k)}) \\ &> (1-\widetilde{\mu}_1)\widetilde{\alpha}_l^{(k)}\nabla^a\varphi_l^{(k)T}d_l^{(k)} + \varepsilon^{(k)},\end{aligned} \tag{3.10}$$

where $\widetilde{y}_l^{(k)} = \widetilde{\alpha}_l^{(k)} d_l^{(k)}$. From P4, we obtain

$$\begin{aligned}&F^a(A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}, \gamma\varepsilon_l^{(k)}) - F^a(x^{(k)}, \varepsilon^{(k)}) \le \\ &F(A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}) - F(x^{(k)}) + \gamma\varepsilon^{(k)}.\end{aligned} \tag{3.11}$$

Using the Mean-Value Theorem, one has

$$F(A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}) - F(x^{(k)}) = [A_l^{(k)T} g(x^{(k)} + \theta A_l^{(k)}\widetilde{y}_l^{(k)})]^T\widetilde{y}_l^{(k)}, \tag{3.12}$$

for some $\theta \in (0, 1)$. It follows from (3.10)-(3.12) that

$$\begin{aligned}[A_l^{(k)T} g(x^{(k)} + \theta A_l^{(k)}\widetilde{y}_l^{(k)})]^T\widetilde{y}_l^{(k)} &> (1-\widetilde{\mu}_1)\nabla^a\varphi_l^{(k)T}\widetilde{y}_l^{(k)} + (1-\gamma)\varepsilon_l^{(k)} \\ &> (1-\widetilde{\mu}_1)\nabla^a\varphi_l^{(k)T}\widetilde{y}_l^{(k)}.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}&[A_l^{(k)T} g(x^{(k)} + \theta A_l^{(k)}\widetilde{y}_l^{(k)}) - A_l^{(k)T} g(x^{(k)})]^T d_l^{(k)} \\ > \;& (1-\widetilde{\mu}_1)\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \nabla\varphi_l^{(k)T}d_l^{(k)} \\ = \;& -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} + (\nabla^a\varphi_l^{(k)} - \nabla\varphi_l^{(k)})^T d_l^{(k)} \\ \ge \;& -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \|A_l^{(k)}\| \cdot \|g^a(x^{(k)}, \varepsilon^{(k)}) - g(x^{(k)})\| \cdot \|d_l^{(k)}\| \\ \ge \;& -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \delta_l\|g^a(x^{(k)}, \varepsilon^{(k)}) - g(x^{(k)})\| \cdot \|d_l^{(k)}\| \\ \ge \;& -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \delta_l\sqrt{2\lambda^{-1}\varepsilon^{(k)}}\|d_l^{(k)}\|,\end{aligned} \tag{3.13}$$

where the third and the last inequality can be obtained in terms of (A3) and P4. Since g is Lipschitzian with constant λ^{-1}, one has

$$\begin{aligned}&[g(x^{(k)} + \theta A_l^{(k)}\widetilde{y}_l^{(k)}) - g(x^{(k)})]^T A_l^{(k)} d_l^{(k)} \\ &\le \theta\lambda^{-1}\widetilde{\alpha}_l^{(k)}\|A_l^{(k)}\|^2 \cdot \|d_l^{(k)}\|^2 \\ &\le \theta\lambda^{-1}\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2.\end{aligned} \tag{3.14}$$

Then, it follows from (3.13) and (3.14) that

$$
\begin{aligned}
\lambda^{-1}\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2 &\geq \theta\lambda^{-1}\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2 \\
&\geq -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \delta_l\sqrt{2\lambda^{-1}\varepsilon^{(k)}}\|d_l^{(k)}\| \\
&\geq -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \delta_l\sqrt{2\lambda^{-1}\beta^{(k)}}\|y_l^{(k)}\|\cdot\|d_l^{(k)}\| \\
&\geq -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)} - \delta_l\alpha_l^{(k)}\sqrt{2\lambda^{-1}\beta^{(k)}}\|d_l^{(k)}\|^2. \qquad (3.15)
\end{aligned}
$$

Since $\widetilde{\alpha}_l^{(k)} = 2\alpha_l^{(k)}$, we have from (3.15) that

$$(\delta_l + \lambda\sqrt{2\lambda^{-1}\beta^{(k)}})\alpha_l^{(k)}\delta_l\lambda^{-1}\|d_l^{(k)}\|^2 \geq -\widetilde{\mu}_1\nabla^a\varphi_l^{(k)T}d_l^{(k)},$$

and since $\beta^{(k)} \to 0$ as $k \to \infty$, the above inequality guarantees the existence of an integer $\overline{k} > 0$ such that for all $k > \overline{k}$

$$\delta_l^2\lambda^{-1}\alpha_l^{(k)} \geq -\widetilde{\mu}_1\|d_l^{(k)}\|^{-2}\nabla^a\varphi_l^{(k)T}d_l^{(k)}$$

or

$$\alpha_l^{(k)} \geq -\lambda\widetilde{\mu}_1\delta_l^{-2}\|d_l^{(k)}\|^{-2}\nabla^a\varphi_l^{(k)T}d_l^{(k)}.$$

We have from (3.2) that

$$\alpha_l^{(k)}\nabla^a\varphi_l^{(k)T}d_l^{(k)} \leq -\lambda\widetilde{\mu}_1\delta_l^{-2}\|d_l^{(k)}\|^{-2}(\nabla^a\varphi_l^{(k)T}d_l^{(k)})^2 \qquad (3.16)$$

and

$$(\nabla^a\varphi_l^{(k)T}d_l^{(k)})^2 \geq \mu_0^2\|\nabla^a\varphi_l^{(k)}\|^2\cdot\|d_l^{(k)}\|^2. \qquad (3.17)$$

Let $\Delta^{(k)a}(y_l^{(k)},\gamma\varepsilon^{(k)}) = \varphi_l^{(k)a}(y_l^{(k)},\gamma\varepsilon^{(k)}) - \varphi_l^{(k)a}(0,\varepsilon^{(k)})$. Combining (3.3), (3.16) and (3.17) and letting $\omega(\lambda,\, \mu_0,\, \widetilde{\mu}_1) = \lambda\mu_0^2\widetilde{\mu}_1$, we obtain

$$
\begin{aligned}
\Delta^{(k)a}(y_l^{(k)},\gamma\varepsilon^{(k)}) &= F^a(A_l^{(k)}y_l^{(k)} + x^{(k)},\gamma\varepsilon_k) - F^a(x^{(k)},\varepsilon^{(k)}) \\
&\leq -\omega(\lambda,\, \mu_0,\, \widetilde{\mu}_1)\delta_l^{-2}\|\nabla^a\varphi_l^{(k)}\|^2 + \varepsilon^{(k)} \\
&\leq -\omega(\lambda,\, \mu_0,\, \widetilde{\mu}_1)(\max\nolimits_{1\leq l\leq p}\delta_l)^{-2}\|\nabla^a\varphi_l^{(k)}\|^2 + \varepsilon^{(k)}.
\end{aligned}
$$

Let $\eta = \omega(\lambda,\, \mu_0,\, \widetilde{\mu}_1)(\max_{1\leq l\leq p}\delta_2)^{-2} > 0$. $\square$

Combining the results given above in Lemmas 3.1–3.6, we present the following PVT-MYR algorithm satisfying the ε-descent condition given above, called the ε-descent PVT-MYR algorithm.

The ε-Descent PVT-MYR Algorithm: for nonsmooth convex minimization (P)

Step 0 `Initialization`

An initial point $x^{(0)} \in R^n$, $B^{(0)} = I_{n\times n}$, $\varepsilon' > 0$, $\varepsilon^{(0)} > 0$, $0 < \gamma < 1$ and set $k = 0$.

Step 1 `Making the Moreau-Yosida regularization`

Step 1a `Find an estimate of the minimizer of Moreau-Yosida regularization`

Given an $\varepsilon^{(k)} > 0$, calculate $p^a(x^{(k)}, \varepsilon^{(k)})$ satisfying (1.8) and (1.9).

Step 1b `Find an estimate of the Moreau-Yosida regularization and gradient`

Formulas (1.10) and (1.11) are used for finding an estimate of the Moreau-Yosida regularization and the corresponding gradient

$$F^a(x^{(k)}, \varepsilon^{(k)}) = f(p^a(x^{(k)}, \varepsilon^{(k)})) + \frac{1}{2}\lambda^{-1}\|p^a(x^{(k)}, \varepsilon^{(k)}) - x^{(k)}\|^2, \tag{3.18}$$

$$g^a(x^{(k)}, \varepsilon^{(k)}) = \lambda^{-1}[x^{(k)} - p^a(x^{(k)}, \varepsilon^{(k)})]. \tag{3.19}$$

Step 1c `Search direction`

Compute a direction $d^{(k)}$ satisfying

$$d^{(k)} = -B^{(k)-1} g^a(x^{(k)}, \varepsilon^{(k)}).$$

Step 2 `Parallelization`

Step 2a `Parallel initialization`

For each $l \in \{1, \ldots, p\}$, choose an $n \times m_l$ matrix $A_l^{(k)}$, such that $d^{(k)} \in \operatorname{span} A_l^{(k)}$.

Step 2b `Choose a direction`

For each $l \in \{1, \ldots, p\}$, choose a direction $d_l^{(k)}$ such that

$$A_l^{(k)} d_l^{(k)} = d^{(k)}.$$

Step 2c `Making the Moreau-Yosida regularization and line search`

The stepsize $\alpha_l^{(k)} > 0$ is chosen according to the Armijo rule

$$\begin{aligned}\varphi_l^{(k)a}(y_l^{(k)}, \gamma\varepsilon^{(k)}) - \varphi_l^{(k)a}(0, \varepsilon^{(k)}) &= F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon_k) - F^a(x^{(k)}, \varepsilon^{(k)}) \\ &\leq \mu_1 \alpha_l^{(k)} [A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})]^T d_l^{(k)} + \varepsilon^{(k)},\end{aligned}$$

where $y_l^{(k)} = \alpha_l^{(k)} d_l^{(k)}$. If $A_l^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)}) \leq \varepsilon'$ for all $l \in \{1, \cdots, p\}$, then stop, otherwise, goto Step 3.

Step 3 `Synchronization`

Choose a vector $x^{(k+1)}$ satisfying

$$F^a(x^{(k+1)}, \gamma\varepsilon^{(k)}) \leq \min_{1 \leq l \leq p} F^a(A_l^{(k)} y_l^{(k)} + x^{(k)}, \gamma\varepsilon^{(k)}).$$

Set $\varepsilon^{(k+1)} = \gamma\varepsilon^{(k)}$.

Step 4 `Update a matrix`

Update $B^{(k)}$ by the BFGS formula

$$B^{(k+1)} = B^{(k)} - \frac{B^{(k)} s^{(k)} s^{(k)^T} B^{(k)}}{s^{(k)T} B^{(k)} s^{(k)}} + \frac{y^{(k)} y^{(k)T}}{s^{(k)T} y^{(k)}},$$

where $s^{(k)} = x^{(k+1)} - x^{(k)}$, $y^{(k)} = g^a(x^{(k+1)}, \varepsilon^{(k+1)}) - g^a(x^{(k)}, \varepsilon^{(k)})$, with $0 < \varepsilon^{(k+1)} < \varepsilon^{(k)}$. Set $k = k + 1$, goto Step 1.

End of the ε-Descent PVT-MYR Algorithm

Note that for each l there exist a sequence of matrices, $\{A_l^{(k)}\}_{k=1}^{\infty}$, satisfying the following conditions

1. $d^{(k)} \in \text{span} A_l^{(k)}$,
2. $A_l^{(k)T} A_l^{(k)}$ is uniformly positive definite,
3. $A_l^{(k)}$ is uniformly bounded.

For example, for each k and l we may choose the matrix $A_l^{(k)} = (d^{(k)}/\|d^{(k)}\|, p_1^{(k)}, \cdots, p_{m_l-1}^{(k)})$, such that columns $p_j \in R^n$, $j = 1, \ldots, m_l - 1$, and $d^{(k)}/\|d^{(k)}\|$ are orthogonal to each other and $\|p_l^{(k)}\| = 1$.

The convergence of the ε-descent PVT-MYR algorithm associated with directions and stepsizes satisfying (3.2) and (3.3) is given by the following theorem.

Theorem 3.1. *Suppose that, at each iteration of the PVT-MYR algorithm, $y_l^{(k)}$ is given by* (3.1) *with $d_l^{(k)}$ and $\alpha_l^{(k)}$ satisfying* (3.2) *and* (3.3), *respectively. Suppose also that the matrices $A_l^{(k)}$ are chosen such that* (A2)–(A4) *are satisfied. Then the sequence $\{x^{(k)}\}$ generated by the PVT-MYR algorithm converges to the unique minimal solution of* (P).

Proof. By virtue of Theorem 2.1 and Lemma 3.6 we have (2.6) and hence the proof is completed. □

4. Rate of convergence. In this section, we investigate the convergence rate of the PVT-MYR algorithm for minimizing a nonsmooth convex function. We assume that the sequence $\{x^{(k)}\}$ generated by the PVT-MYR algorithm is convergent to the minimizer of f, i. e.,

$$\lim_{k\to\infty} \|x - x^*\| = 0,$$

where $x^* = \arg\min_{x\in R^n} f(x)$. The following two conditions are used for studying the rate of convergence

(B1) $\|g(x^{(k)})\|^2 - \|g^a(x^{(k)}, \varepsilon^{(k)})\|^2 \le -2\varepsilon^{(k)}/(\eta\beta(\lambda p)^{-1})$,

(B2) $c_3 < \eta\beta p^{-1}$, where c_3 is a positive constant.

We now give the result on the convergence rate of the PVT-MYR algorithm.

Theorem 4.1. *Let $\{x^{(k)}\}$ be a sequence generated by the PVT-MYR algorithm under the following assumptions*

a. *The objective function f is strongly convex, satisfying* (B1) *and* (B2)*;*

b. $y_l^{(k)}$, $l = 1, \cdots, p$, $k = 1, \cdots,$ *are chosen such that* (2.6) *in Step* 1 *is satisfied;*

c. $z^{(k)}$, $l = 1, \cdots, p$, $k = 1, \cdots,$ *are chosen such that* (2.7) *in Step* 2 *is satisfied;*

d. *Matrices $A_l^{(k)}$ satisfy* (A2) *and* (A3).

Then $\|x^{(k)} - x^*\|$ *converges R-linearly to zero.*

Proof. From (2.10), we have

$$\begin{aligned} F(x^{(k+1)}) - F(x^{(k)}) &\le -\eta p^{-1}\Big(\sum \|\nabla^a \varphi_l^{(k)}\|^2\Big) + 2\varepsilon^{(k)} \\ &= -\eta p^{-1}\|A^{(k)T} g^a(x^{(k)}, \varepsilon^{(k)})\|^2 + 2\varepsilon^{(k)} \\ &\le -\eta\beta p^{-1}\|g^a(x^{(k)}, \varepsilon^{(k)})\|^2 + 2\varepsilon^{(k)}. \end{aligned}$$

We have from (B1) and (B2) that

$$\|g(x^{(k)})\|^2 - \|g^a(x^{(k)}, \varepsilon^{(k)})\|^2 \le -2\varepsilon^{(k)}/(\eta\beta p^{-1}),$$

and

$$-\eta\beta p^{-1}\|g^a(x^{(k)}, \varepsilon^{(k)})\|^2 + c_3\|g(x^{(k)})\|^2 + 2\varepsilon^{(k)} \le 0.$$

Thus, one has

$$F(x^{(k+1)}) - F(x^{(k)}) \le -c_3\|g(x^{(k)})\|^2. \tag{4.1}$$

Combining (4.1), P1 and P2, one has

$$F(x^{(k)}) - F(x^{(k+1)}) \ge c_3 c_2^2 c_1^{-1}(F(x^{(k)}) - F(x^*)).$$

This in turn implies that

$$F(x^{(k+1)}) - F(x^*) \le c(F(x^{(k)}) - F(x^*)), \tag{4.2}$$

where $c = 1 - c_3 c_2^2 c_1^{-1} \in (0,1)$. Thus, $\{F(x^{(k)})\}$ converges Q-linearly to $F(x^*) = f^*$.

Since (4.1) and (4.2) imply that

$$F(x^{(k)}) - F(x^*) \ge (1-c)\|g(x^{(k)})\|^2$$

and

$$F(x^{(k)}) - F(x^*) \le c^k(F(x^{(0)}) - F(x^*)),$$

respectively, we obtain

$$\|g(x^{(k)})\|^2 \le c^k(1-c)^{-1}(F(x^{(0)}) - F(x^*)).$$

By P2, we have that $\|x^{(k)} - x^*\|$ converges R-linearly to zero. □

5. Appendix. The proximal solution problem (1.2) can be computed by the algorithm proposed in [3], stated below.

Proximal Point Algorithm: for nonsmooth minimization problems

$$p(x^{(n)}) = \arg\min\{f(z) + \frac{1}{2t_n}\|y - x^{(n)}\|\}.$$

Step 0 Fix, for example, $k > 1$ and $m \in (0, 1)$. Start from $x^{(1)} \in R^n$, set $n = 1$.

Step 1 Set $k = 1$, start from some $y^{(k)} = y^{(1)}$.

Step 2 Set

$$\varepsilon = k[f(x^{(n)}) - f(y^{(k)} - \frac{m}{t^{(n)}}\|y^{(k)} - x^{(n)}\|^2)].$$

If

$$\frac{x^{(n)} - y^{(k)}}{t_n} \in \partial_\varepsilon f(x^{(n)}),$$

then goto Step 3; otherwise compute $y^{(k+1)}$, increase k by 1 and execute Step 2 again.

Step 3 Set $x^{(n+1)} = y^{(k)}$, increase n by 1 and loop to Step 1.

End of the Proximal Point Algorithm

Acknowledgements. The authors are extremely indebted to the anonymous referees for valuable and constructive comments and appreciate Profs. Vesselin Drensky and Roger Van Keer, and Dr. Nikolay Kirov for providing an opportunity for the authors to publish their paper on this Journal. This work was supported by the State Foundations for Ph. D. Units (2003-05), NSF of China (10001007) and Research Foundation of DUT (2002-03).

REFERENCES

[1] A. Auslender. Numerical methods for nondifferentiable convex optimization. *Math. Program. Study* **30** (1987), 102–126.

[2] D. P. Bertsekas, J. N. Tsitsiklis. Parallel and Distributed Computation: Numerical Methods. Prentice-Hall: Englewood Cliffs, New Jersey, 1989.

[3] R. Correa, C. Lemaréchal. Convergence of some algorithm for convex minimization. *Math. Program.* **6** (1993), 261–273.

[4] C. C. Chou, K. F. Ng, J. S. Pang. Minimizing and stationary sequences of constrained optimization problems. *SIAM J. Control Optim.* **36** (1998), 1908–1936.

[5] M. Fukushima. Parallel variable transformation in unconstrained optimization. *SIAM J. Optim.* **8** (1998), 658–672.

[6] M. Fukushima. A descent algorithm for nonsmooth convex optimization. *Math. Program.* **30** (1984), 163–175.

[7] M. Fukushima, L. Qi. A globally and superlinearly convergent algorithm for nonsmooth convex minimization. *SIAM J. Optim.* **6** (1996), 1106–1120.

[8] M. C. Ferris, O. L. Mangasarian. Parallel variable distribution. *SIAM J. Optim.* **4** (1994), 102–126.

[9] S. P. Han. Optimization by updated conjugate subspaces. In: Numerical Analysis: Pitman Research Notes Mathematics Series 140, (Eds D. F. Griffiths, G. A. Watson) Longman Scientific and Technical, Burnt Mill, England, **5** (1986), 82–97.

[10] S. P. Han, G. Lou. A parallel algorithm for a class of convex programs. *SIAM J. Control Optim.* **26** (1988), 346–355.

[11] J. Hiriart-Urruty, C. Lemaréchal. Convex Analysis and Minimization Algorithms. Springer Verlag, Berlin, 1993.

[12] C. S. Liu, C. H. Tseng. Parallel synchronous and asynchronous space-decomposition algorithms for large-scale minimization. *Comput. Optim. Appl.* **17** (2000), 85–107.

[13] O. L. Mangasarian. Parallel gradient distribution in unconstrained optimization. *SIAM J. Control Optim.* **33** (1995), 1916–1925.

[14] J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93** (1965), 273–299.

[15] J. M. Ortega, W. C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, 1970.

[16] A. I. Rauf, M. Fukushima. Globally convergent BFGS method for nonsmooth convex optimization. *J. Optim. Theory Appl.* **104** (2000), 539–58.

[17] M. V. Solodov. New inexact parallel variable distribution algorithms. *Comput. Optim. Appl.* **7** (1997), 165–182.

[18] Z. Wei, L. Qi. Convergence analysis of a proximal newton method. *Numer. Funct. Anal. Optim.* **17** (1996), 463-472.

[19] E. Yamakawa, M. Fukushima. Testing parallel variable transformation. *Comput. Optim. Appl.* **13** (1999), 1-22.

[20] K. Yosida. Functional Analysis. Springer Verlag, Berlin, 1964.

Li-Ping Pang
CORA, Department of Applied Mathematics
Dalian University of Technology
Dalian 116024, China
e-mail: lipingpang@163.com

Zun-Quan Xia
CORA, Department of Applied Mathematics
Dalian University of Technology
Dalian 116024, China
e-mail: zqxiazhh@dlut.edu.cn,
zqxia@mail.dlptt.ln.cn.

Received July 23, 2002
Revised October 21, 2002